51

Reprinted from JOURNAL OF EXPERIMENTAL PSYCHOLOGY,
Vol. 45, No. 4, April, 1953
Printed in U. S. A.

STUDIES OF DISTRIBUTED PRACTICE: X. THE INFLUENCE OF INTRALIST SIMILARITY ON LEARNING AND RETENTION OF SERIAL ADJECTIVE LISTS

BENTON J. UNDERWOOD[1]

Northwestern University[2]

The major purpose of this study was to discover the influence of intralist similarity and intertrial interval on retention of serial adjectives. With regard to retention as a function of intralist similarity, several studies (4, 6, 7, 8) indicate that while intralist similarity may influence rate of learning, it has no influence on recall taken 24 hr. following learning. As yet, however, no tests of retention have been made for serial adjectives of varying levels of intralist similarity.

The data are conflicting concerning retention of rote-learned materials following acquisition with varying intertrial intervals. With serially learned nonsense syllables, Hovland (3) found better retention after spaced practice than after massed practice. We have found quite the opposite with comparable materials (6). Retention of paired adjectives may (4) or may not (8) be enhanced by distributed practice depending, we believe, upon degree of learning before the retention interval. Although degree of learning has not clearly been shown to be the critical variable, some evidence (summarized in 8) suggests that with high degrees of learning massed practice gives better recall, and with low degrees of learning spaced practice gives better recall. In any event, data available at the present time do not justify generalizations concerning retention following learning by massed and distributed practice. Since there are no results on this variable for serial adjectives, the present experiments will fill this gap.

Previous experiments (5, 9) on learning of serial adjectives as a function of massed vs. distributed practice have shown small but consistent differences in favor of faster learning with distribution. The present experiments will give additional data on this relationship, for obviously, in order to measure retention following different intertrial intervals, learning rates under these intervals will also be available.

PROCEDURE: EXP. I, II, AND III

General.—The first data to be reported are based on three experiments. Two additional experiments make up part of the report but will not be described until after the results of the first three studies are given. The first three experiments are differentiated only by degree of intralist similarity of the adjective lists. In Exp. I, intralist similarity is low; in Exp. II, medium; and in Exp. III, intralist similarity is high. Each experiment had three different conditions, these conditions varying in length of intertrial rest interval. These intervals were 2 sec. (massed), 30 sec., and 60 sec.

Lists.—Since there were three conditions in each experiment, and since a given *S* served in all three conditions, three different experimental lists were required for each experiment. Each serial list consisted of 14 two-syllable adjectives, but since the first word was used only as an anticipatory cue, *S* learned only 13 items. All adjectives were taken from Haagen (2). In constructing the lists, the 14 items to be used were formed into five sets in which four sets consisted of three words and one set of two words. In varying similarity, synonymity of items within

[1] R. L. Morgan and E. J. Archer supervised the gathering of the data; Mr. Archer and Jack Richardson are largely responsible for the statistical analysis.

[2] This work was done under Contract N7onr-45008, Project NR 154-057, between Northwestern University and the Office of Naval Research.

sets was varied. That is, for low similarity the items within sets making up the lists would have low synonymity while for high similarity the items within sets had a high-synonymity rating. The lists, and sets within lists, are illustrated below for each level of similarity. Sets are enclosed in parentheses.

Exp. I: Low Similarity: (fickle, heedless, fitful) (urgent, crying, required) (sickly, bedfast, feeble) (complete, perfect, utter) (empty, yawning)

Exp. II: Medium Similarity: (anxious, distressed, troubled) (cautious, discreet, guarded) (swollen, bloated, puffy) (liquid, flowing, solvent) (spoken, talking)

Exp. III: High Similarity: (angry, enraged, wrathful) (complete, entire, total) (double, dual, twofold) (royal, regal, kingly) (liquid, fluid)

If one studies the above illustrations, it will be recognized that degree of similarity within sets increases from Exp. I through Exp. II. The method of increasing similarity, it will be realized, does not increase the number of potential interfering tendencies but increases the strength of those tendencies. In Haagen's scaling, .5 indicates highest similarity and 6.5 lowest. Averaging the synonymity ratings for the sets of each list for Exp. I gives means of 4.04, 3.80, and 4.01; for Exp. II the means are 2.40, 2.42, and 2.40; for Exp. III the means are 1.22, 1.31, and 1.22.

In ordering items for a list, of course, synonyms were never placed together but were scattered throughout the list. The individual familiarity ratings of all adjectives in a list averaged about the same for all lists for all three levels of similarity. Obvious associative cues (e.g., same prefix on two successive words) were kept at a minimum.

A single practice list, learned by all *Ss*, was of medium similarity. All lists were presented on Hull-type drums at a 2-sec. rate, with the anticipation method of learning used throughout.

Specific conditions.—On the practice day *S* learned the practice list to seven correct responses on a single trial by massed practice (2 sec. between trials). Instructions for symbol cancellation, used to fill rest intervals between trials, were followed by continued learning of the list to one perfect trial with 30-sec. rest between each trial. After 5-min. rest the practice list was recalled and relearned.

On the experimental days *S* learned a list to one perfect recitation under the intertrial interval appropriate for the day. After 24 hr. the list was recalled and relearned by massed practice. On the first three experimental days *S* learned a new list each day which was in turn recalled the following day. On the last of the four experimental days only recall and relearning of the list learned the previous day were required. Thus, *S* served in five sessions, the first being a practice session. The four experimental sessions came on four successive days in order to meet the 24-hr. retention interval for each list.

A total of 36 undergraduate students was used in each experiment. Conditions of intertrial rest and lists were completely counterbalanced against practice effects. The method for statistical analysis of such designs has been reported elsewhere by Archer (1).

RESULTS: EXP. I, II, AND III

Practice list.—The mean number of trials to learn the practice list to one perfect recitation was 27.36, 29.72, and 26.25 for Exp. I, II, and III, respectively. *F* is less than one. The product-moment correlation between trials to learn practice list and trials to learn all three experimental lists for all three experiments combined is .60±.10. The mean number of errors per trial in learning the practice list was 1.83, 1.54, and 1.74. *F* is 1.32, with an *F* of 3.09 needed for significance at the .05 level of confidence. The correlation between errors

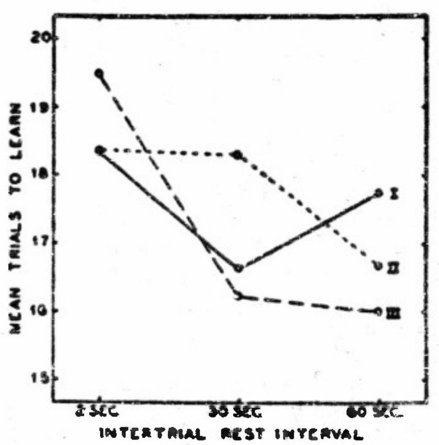

FIG. 1. Learning as a function of intertrial interval and intralist similarity. Exp. I had lists of low similarity; Exp. II medium; and Exp. III, high intralist similarity.

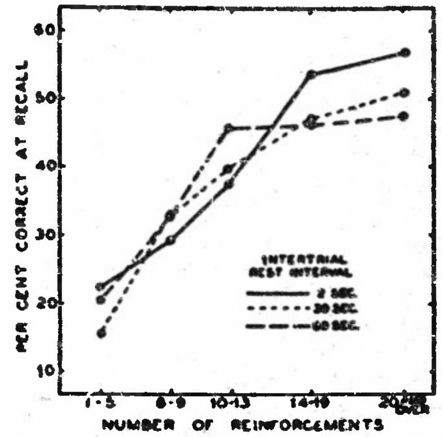

FIG. 2. Retention as a function of intertrial rest with number of reinforcements (correct responses) during learning held constant

on the practice day and errors on experimental days was .64 ± .10. The three groups of *Ss* may be considered comparable.

Experimental lists.—The mean number of trials to learn the lists of varying similarity under three intertrial rest conditions is shown in Fig. 1. The statistical analysis of the distributions on which these means are based has shown: (*a*) Similarity as varied here is not an effective variable, *F* being less than one. (*b*) Intertrial rest is a significant source of variance. The *F* for intertrial rest is 4.48. With 2 and 200 *df*, *F* is 3.04 at .05 level of confidence, and 4.71 at the .01 level. (*c*) Interaction between intertrial rest and intralist similarity is far from significant (*F* is 1.33).

Mean errors per trial did not vary as a function of intertrial interval. Combining all three intertrial rest conditions for each experiment, the mean number of errors per trial was 1.22, 1.21, and 1.45, for Exp. I, II, and III, respectively. The distributions on which these means are based do not differ significantly.

Recall and relearning.—Analysis of the 24-hr. recall shows no significant source of variance. The trend for the raw recall scores was for better recall for high-similarity lists than for low-similarity lists. In order to evaluate the influence of intertrial rest on retention we have made an item analysis of original learning by grouping items having comparable number of reinforcements (correct anticipations) and then determining percentage correct at recall for each grouping. The result of this analysis is shown in Fig. 2. While the differences are small, it may be noted that recall is better following massed practice than following spaced practice for items receiving a large number of reinforcements. The situation is reversed for items with a moderate number of reinforcements. As summarized elsewhere (8), comparable trends have been noted in several other studies.

Neither interlist similarity nor intertrial interval was effective during relearning.

All findings for learning, recall, and relearning indicate that similarity as manipulated here was an ineffective variable. The only clear positive finding of the three experiments was that distributed practice produced faster learning than massed practice. Even here the level of significance was not high and confirms previous findings (5, 9) that with serial adjectives distributed practice produces small but consistent amounts of facilitation.

In view of the fact that similarity did not influence rate of learning, we have no adequate test of retention as a function of intralist similarity. Therefore, we have run two additional experiments to obtain more definitive data on this matter. The purpose was to obtain lists varying in similarity which would produce differences in rate of learning.

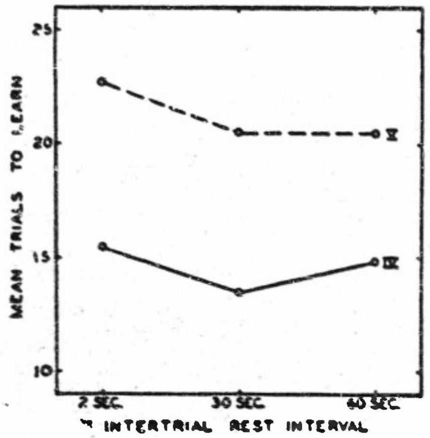

FIG. 3. Learning as a function of intertrial interval and interlist similarity. Experiment IV had very low intralist similarity; Exp. V had very high intralist similarity.

PROCEDURE: EXP. IV AND V

The only difference between these two experiments and those reported above was in lists. In Exp. IV the lists had very low similarity. No known meaningful similarity existed between items within a list and all apparent instances of formal similarity were eliminated. In Exp. V, on the other hand, lists were constructed to have high similarity by increasing the number of synonymous words within a list. The 14-item lists were made with three sets of highly synonymous adjectives, one set containing four synonyms and two sets containing five. This is to be contrasted with Exp. III where we used four sets of three synonyms and one set of two. In short, all details of the experiments were the same as those reported above except that Exp. IV had lists constructed to have lower intralist similarity than Exp. I, and Exp. V had lists constructed to have higher intralist similarity than Exp. III.

RESULTS: EXP. IV AND V

Practice list.—The mean number of trials required to learn the practice list was 28.61 for Exp. IV, and 27.31 for Exp. V. The corresponding mean number of errors per trial was 1.36 and 1.40. Since these means for either measure do not differ significantly, we may consider the two groups of *Ss* equivalent.

Experimental lists.—The mean number of trials required to learn the lists are plotted in Fig. 3. It is apparent that similarity produced wide differences in rate of learning. Furthermore, as in the first three experiments reported in this paper, distributed conditions produced somewhat more rapid learning than did massed conditions. There is no evidence for interaction between intertrial rest and intralist similarity.

For all three conditions, more errors were made in learning high-similarity lists than in learning low-similarity lists but these differences were not great enough to be reliable statistically.

Retention.—Intertrial interval during learning produced no significant differences in recall. However, differences in recall attributable to similarity were very large, with the better retention occurring for the high-similarity lists. For all three conditions combined, the mean number of items recalled from low-similarity lists was 3.30; for the high-similarity items the corresponding value was 5.00. The F

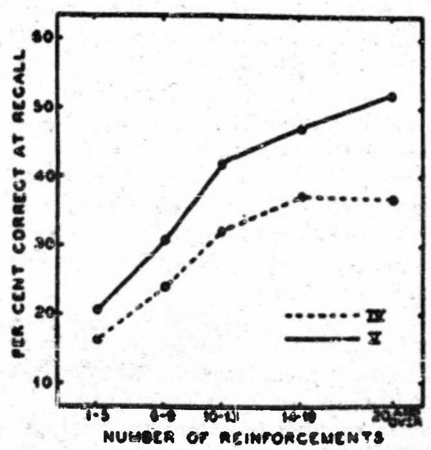

FIG. 4. [illegible]tion as a function of intralist similarity with number of reinforcements during learning held constant. Experiment IV had low similarity; Exp. V, high.

is 11.35, with an F of 7.01 needed for the .01 level with 1 and 70 *df*.

These retention results as a function of similarity must be evaluated cautiously. Since the number of trials required to learn the two sets of lists was widely different, it is apparent that mean frequency of reinforcements for items in the two lists was different. Therefore, differences at recall may reflect this variable and not similarity. Accordingly, an item analysis has been made in which number of reinforcements during learning is held constant for the two lists and percentage correct at recall determined. The results of this analysis are shown in Fig. 4. It can be seen that even with number of reinforcements held constant, recall of high-similarity items is superior to recall of low-similarity items. To make sure that Fig. 4 is not a result of bias in grouping reinforcements for learning, we have also analyzed recall for number of reinforcements from 1 through 20 separately. At nearly all points recall was better for high-similarity lists than for low-similarity lists. Clearly, the conclusion is that retention of serial adjectives with high intralist similarity is better than retention of serial adjectives of low intralist similarity.

Relearning required longer for high-similarity lists than for low, and more errors per trial were made in relearning the high-similarity lists than the low, but neither trend had statistical reliability.

DISCUSSION

As in previous studies (5, 9) we have found that learning of serial adjectives is facilitated by distributed practice but this facilitation is not great. In all studies the level of statistical significance has not been high but always learning is faster with distribution than with massing. Furthermore, we have no clear evidence in any of these studies that there is interaction between intertrial interval and interlist similarity. As in the case of nonsense syllables (6) there is no basis for believing that difficulty as manipulated by intralist similarity is an important variable in the study of massed vs. distributed practice.

Retention as measured by recall scores has not varied as a function of intertrial interval in the five experiments reported here. Consistent with previous findings (8), however, is the trend noted from the item analyses of recall that heavily reinforced items are better recalled following massed practice than following distributed practice. Conversely, if the item is reinforced only a few times during learning, it will be better retained if learned under distributed practice than if learned under massed practice. It should be emphasized that these are only *trends* noted in the data reported here and in several previous experiments. To verify clearly these suggested principles, independent experiments will be necessary in which degree of learning is manipulated.

Finally, it should be pointed out that in all of our experiments on massing vs. distribution, whether using serial learning or paired-associate learning, or whether using meaningful material or nonsense syllables, no evidence has been found that intralist similarity is an important variable influencing recall. In the present study (Exp. IV and V) with degree of learning held constant, recall was consistently better for lists of high similarity than for those of low similarity. We do not believe that this finding is a contradiction to the above general statement. Rather, we believe that it is a reflection of the particular technique used to vary

similarity and as such represents an exception to the more general findings of previous experiments. It is, of course, important to know that there are ways in which similarity can be manipulated so that recall will be directly correlated with similarity, but the bulk of the evidence, in which similarity has been varied in many different ways, indicates the present finding to be an exception. In our high-similarity lists in the present experiments there were three clusters of synonyms. In constructing the lists two words from a given cluster were never allowed to be serially contiguous. With such a system it is possible to see how recall could be facilitated. Thus *S*, having just seen one word of a given cluster, may learn that the next word cannot be from the same cluster. Or, having seen two or three words from a given cluster during the first part of a trial, *S* might know with some certainty that among the last few words there would be one or two more words from that cluster.

In short, there are ways by which *S* could increase his recall for high-similarity lists but this increment could well be relevant only to the particular lists used here because of the method used in constructing them. Therefore, we do not feel that generalizations should be made from the present finding that high-similarity lists are recalled better than low-similarity lists.

Summary

Three experiments were performed to determine the influence of intralist similarity and intertrial interval on learning and retention of serial adjective lists. Similarity was manipulated by varying degree of synonymity of clusters within lists. Three intertrial rests of 2, 30, and 60 sec. were used for each experiment. The 36 *Ss* in each experiment learned each list to one perfect trial following which retention was measured after 24 hr.

The results showed that the method of manipulating similarity did not produce differences in rate of learning. Learning was more rapid for the 30- and 60-sec. intertrial rest conditions than for the 2-sec. condition, thus confirming previous findings. No differences in retention were evident for either variable.

Two additional experiments were performed in an attempt to produce differences in rate of learning as a function of similarity. By manipulating number of items with a cluster of synonyms, wide differences in rate of learning were achieved. Again, distributed practice facilitated learning. The high-similarity lists were better recalled than low-similarity lists. Since several previous studies, using different materials and learning methods, and different techniques of manipulating similarity have shown little difference in recall as a function of interlist similarity, it was concluded that the better recall of the high-similarity lists in the present experiments was a function of the particular method of producing high similarity.

(Received September 2, 1952)

References

1. Archer, E. J. Some greco-latin analysis of variance designs for learning studies. *Psychol. Bull.*, 1952, 49, 521–537.
2. Haagen, C. H. Synonymity, vividness, familiarity, and association-value ratings of 400 pairs of common adjectives. *J. Psychol.*, 1949, 30, 185–200.
3. Hovland, C. I. Experimental studies in rote-learning theory: VI. Comparison of retention following learning to same criterion by massed and distributed practice. *J. exp. Psychol.*, 1940, 26, 568–587.

4. Underwood, B. J. Studies of distributed practice: II. Learning and retention of paired-adjective lists with two levels of intra-list similarity. *J. exp. Psychol.*, 1951, 42, 153–161.
5. Underwood, B. J. Studies of distributed practice: III. The influence of stage of practice in serial learning. *J. exp. Psychol.*, 1951, 42, 291–295.
6. Underwood, B. J. Studies of distributed practice: VII. Learning and retention of serial nonsense lists as a function of intra-list similarity. *J. exp. Psychol.*, 1952, 44, 80–87.
7. Underwood, B. J. Studies of distributed practice: VIII. Learning and retention of paired nonsense syllables as a function of intralist similarity. *J. exp. Psychol.*, 1953, 45, 133–142.
8. Underwood, B. J. Studies of distributed practice: IX. Learning and retention of paired adjective lists as a function of intralist similarity. *J. exp. Psychol.*, 1953, 45, 143–149.
9. Underwood, B. J., & Goad, D. Studies of distributed practice: I. The influence of intra-list similarity in serial learning. *J. exp. Psychol.*, 1951, 42, 125–134.